



ICiS

Institute of Computing in Science

Methods and Terabase Metagenomics

Summer Session 1:

July 17 to July 24, 2010

In Snowbird, Utah

Dionysios Antonopoulos, PhD

Assistant Biologist, Microbial Systems



Dr. Antonopoulos joined the IGSB and Biosciences Division at Argonne National Laboratory in June 2008 and holds a joint appointment in the Department of Medicine, Section of Gastroenterology, at the University of Chicago.

His present research is focused on soil metagenomics using next-generation DNA sequencing technologies, as well as applying the same analytical tools to understanding a variety of diseases of the GI tract.



Institute for
Genomics &
Systems Biology

Pavan Balaji

- Position:
 - Assistant Computer Scientist, Argonne National Laboratory
- Research Background:
 - Programming models and runtime systems for very large-scale systems
 - Massive parallelism and concurrency support
- Research Interests:
 - Application domain scalability (computational biology, chemistry, nuclear physics) to post-petascale and exascale computing
 - Communication and I/O



Web: <http://www.mcs.anl.gov/~balaji>

Email: balaji@mcs.anl.gov

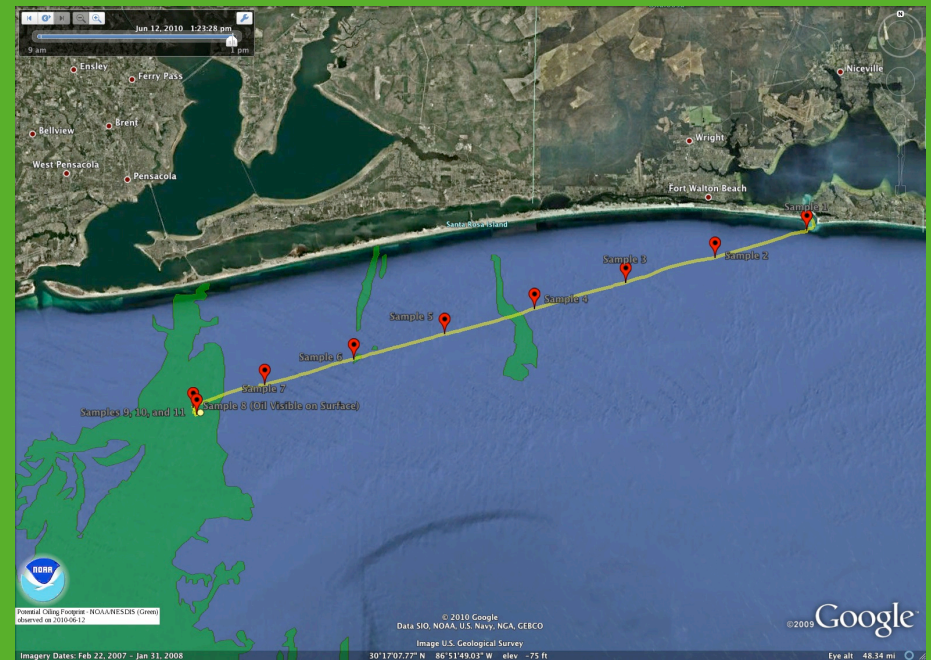


Chris Brown
Undergraduate Researcher
The University of Florida
Department of Microbiology and Cell Science



I am an undergraduate at the University of Florida majoring in Microbiology and Cell Science with a minor in English. I work with Dr. Eric Triplett on metagenome projects involving Type 1 diabetes and the Gulf of Mexico (post Deepwater Horizon oil spill).

My interests are in finding the most reliable methods for annotating metagenomes with functional and taxonomic information to compare microbial diversity between treatments.



On 12 June 2010 my lab sampled surface water off of the coast of Destin Harbor, FL as oil from the Deepwater Horizon oil spill began to land on Florida beaches. We are currently conducting 16S rRNA sequencing on these samples with future plans for metagenomic sequencing.

C. Titus Brown / Michigan State University

Assistant Professor - ctb@msu.edu - <http://ged.msu.edu/>
Computer Science / Microbiology and Molecular Genetics;
BEACON NSF Center on Evolution In Action



Interests

(Open, flexible, scalable, reusable) computational infrastructure for moving from sequence to function.

Estimating total community depth and sequencing error rates w/o a reference.

Assembly and pre-assembly filtering.

“Wow, that’s a lot of data” issues.

Cloud computing approaches.

Narayan Desai

desai@mcs.anl.gov

- HPC System Software
- System Management Software
- Experimental Hardware
- Scaling
Metagenomics
Pipeline Throughput



Jonathan Eisen

Dirk Evers

Director, Computational Biology, Illumina

- PhD Bielefeld University
 - RNA secondary structure prediction
- Exelixis Deutschland
 - forward & backward genetic screens in Zebrafish
- Executive Director, International Graduate School in Bioinformatics and Genome Research
 - high performance biocomputing, de novo assembly, suffix trees, index structures
- Illumina
- Research Interests
 - Human Genomics
 - Cancer Genomics
 - Metagenomics
 - De Novo Assembly



Wu Feng



Associate Professor, Computer Science
and Electrical & Computer Engineering
Co-Director, NSF Center for High-
Performance Reconfigurable Computing

Overarching Goal

- Application of high-performance computing (HPC) to the sciences & engineering



HPC + Bio Projects

- mpiBLAST (<http://www.mpiblast.org/>)
 - Applications of mpiBLAST: Pathogen detection pathogen signature design, 'finding missing genes' in genomes
 - Sequence Analysis (Pairwise & Multiple)
 - Short-Read Mapping
 - Molecular Dynamics
 - Neuroinformatics
-
- Recent Infrastructure Award for Bio
 - \$2M multi-scale supercomputer: desktop supercomputing to datacenter supercomputing

Recent "Accomplishments"

New Software Could Smooth Supercomputing Speed Bumps (<http://www.scientificamerican.com/article.cfm?id=opencl-smooths-supercomputing>);

Sharing the Wealth of Data (<http://www.saworldview.com/article/sharing-the-wealth-of-data>); Harnessing Power from 7 Supercomputers, mpiBLAST Finds

Undetected Genes (<http://www.scientificcomputing.com/news-HPC-Harnessing-Power-from-7-Supercomputers-mpiBLAST-Finds-Undetected-Genes-041510.aspx>).



*Have mountain?
Will scale ...*



Research Group

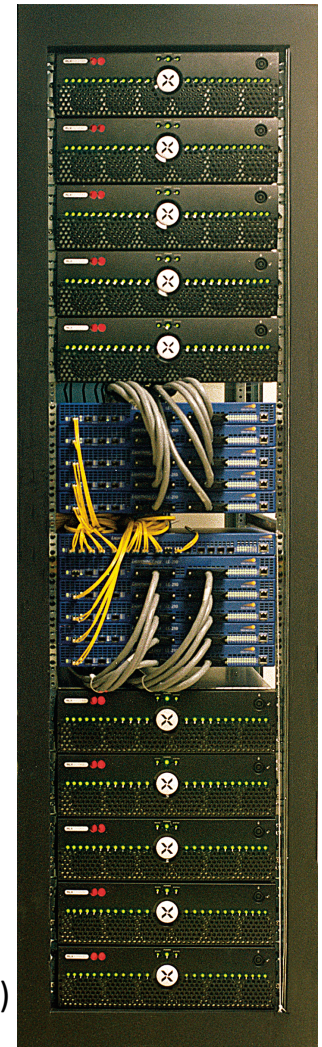
- Synergy Lab
<http://synergy.cs.vt.edu/>

Background

- *Industry & Lab*: LANL('98-'06), Orion Multisystems ('04-'05), Vosaic ('97), NASA Ames ('93), IBM T.J. Watson Rsch ('90).
- *Academia*: U. Illinois ('96-'98), Purdue U. ('98-'00), Ohio St. U. ('00-'03), Virginia Tech ('06-now)

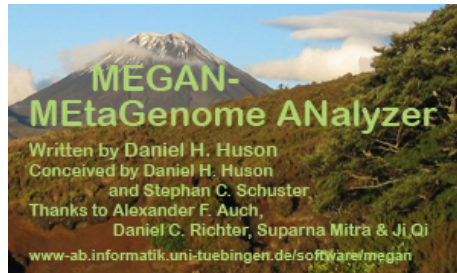
Education

- B.S., Computer Engg. & Music, Penn St. U. 1988.
- M.S., Computer Engg., Penn St. U., 1990.
- Ph.D., Computer Science, U. Illinois, 1996.



Jack Gilbert

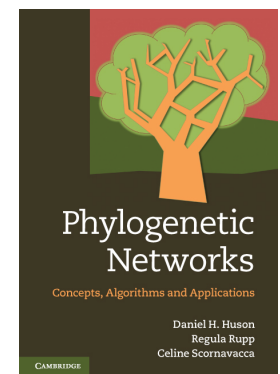




Daniel Huson
Professor of Algorithms in Bioinformatics
Center for Bioinformatics, Tübingen



- Work on methods and algorithms for
 - Metagenome analysis and comparison
 - Phylogenetic analysis using networks
- Software
 - MEGAN – metagenome analyzer
 - MetaSim, metagenome simulator
 - SplitsTree, Dendroscope – trees & networks
- Current projects
 - Metagenome analysis of marine data, mummy data, medical data





Janet Jansson Metagenome Projects

TWINS WITH CROHN'S DISEASE:

12 samples, 454 Titanium. Ongoing Collaboration with Claire Fraser-Liggett. Plus proteomics (ORNL)



DEEPWATER HORIZON OIL SPILL:

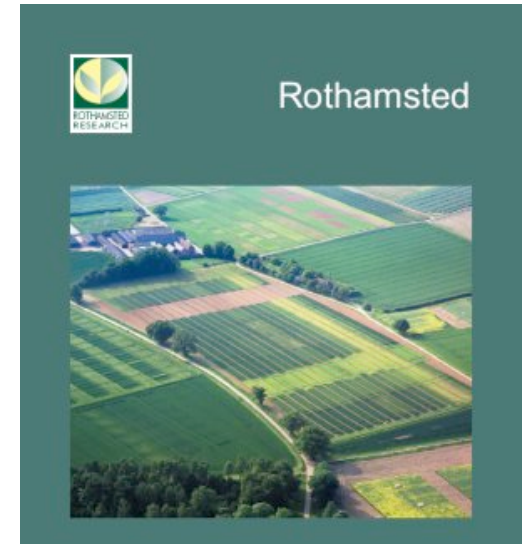
3 samples. Illumina and 454 Titanium (ongoing) Plus single cell genomics



DEEPSOIL: Rothamsted, UK Bare Fallow & Grassland Illumina (2 samples, ~ 50 Gigabases completed) Collaboration with Folker Meyer, Dirk Evers & others

GREAT PRAIRIE & adjacent cultivated corn soils: Wisconsin, Iowa & Kansas Illumina & 454 Titanium (8 samples > 60 Gigabases completed) Plus pyrotags Collaboration with Jim Tiedje & colleagues at JGI

ALASKAN PERMAFROST & Active layer: Illumina (12 samples ~ 40 Gigabases completed) Plus pyrotags Collaboration with Eddy Rubin & Rachel Mackelprang (JGI) & Mark Waldrop (USGS)



- Dr. James Knight

- R&D Fellow at 454 Life Sciences

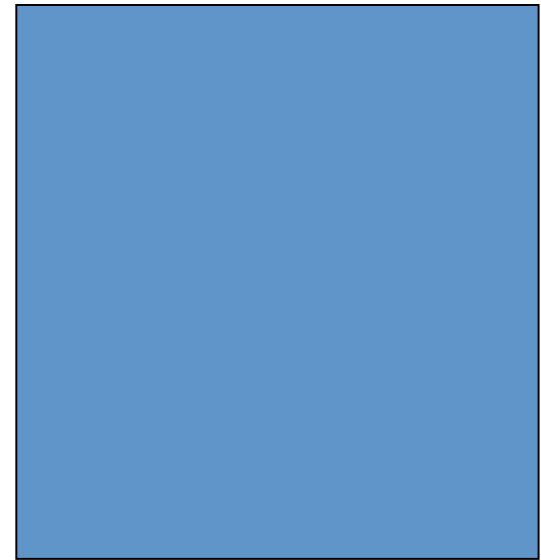
- Author of the Newbler assembler

- Background

- B.A., M.S., Ph.D. in Computer Science
 - Computational biology algorithm development
 - Drug discovery analysis software work
 - Gene expression, protein-protein interaction, ...

- Interests

- Assembly, variation detection, metagenomics
 - How can you really organize genomic/genetic information?





Rob Knight

HHMI/University of
Colorado at Boulder

Background:

BSc: Biochemistry, Otago, 1996

PhD: Ecology & Evolution,
Princeton, 2001; genetic code

Postdoc: Molecular Biology, CU,
2001-2004; SELEX

Interests:

Microbial community analysis

Molecular evolution

Multivariate statistics

High-performance computing

Human microbiome

Eugene Kolker

1. Omics Data Analysis & Integration

2. Discovery Proteomics & Protein Annotation

3. Health Informatics & Predictive Analytics

4. Data-Intensive Research



The Environmental Genomics Lab @ GaTech

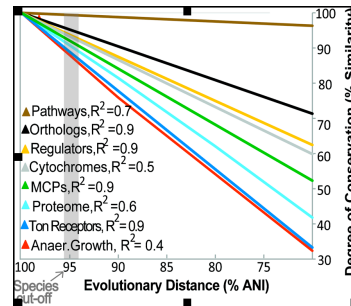
At the interface of microbiol ecology with computers & engineering

Interests

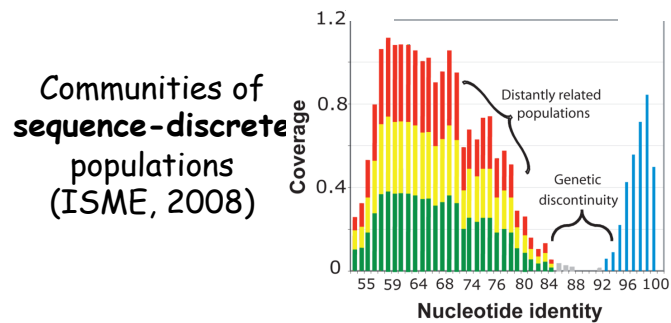
The group

1. Comparative genomics to understand bacterial speciation and genome evolution.

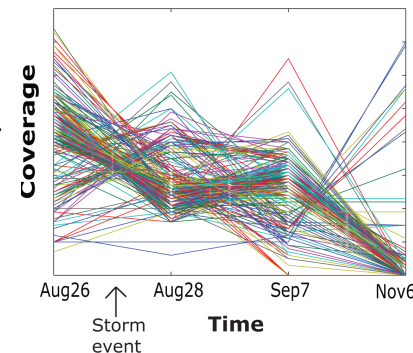
Genotype vs. phenotype in *Shewanella*
(PNAS, 2009)



2. Metagenomics to understand community complexity & diversity. Biotech applications of diversity.



Time series of lake Lanier microbial communities



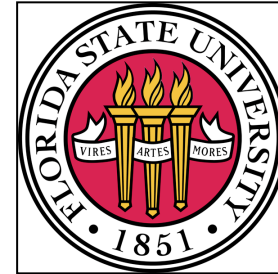
More info: <http://enve-omics.gatech.edu/>





Joel Kostka

Dept. of Earth, Ocean, and Atmospheric Science
Florida State University



- Microbial ecology, biogeochemistry
- Role of microorganisms in ecosystem function
- Marine and aquatic environments; soils/sediments; subsurface
- Microbiology Team Leader, Oak Ridge Integrated Field Research Challenge, DOE-BER
- Nitrogen removal by denitrification and anammox
- Microbial oil degradation in beach sands

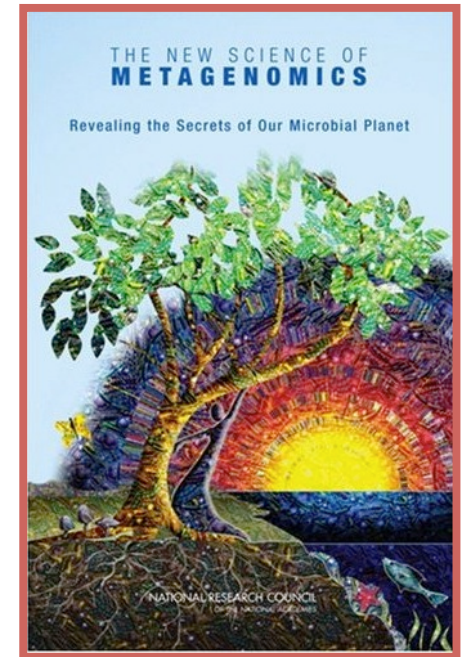
Nikos



Metagenomics



Rachel Mackelprang



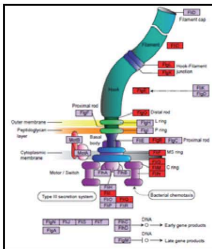
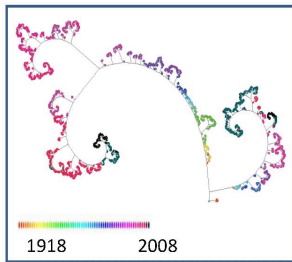
Extract all DNA from
permafrost samples

(before and after thawing)

Sequence
community DNA
(both phylogenetic
and functional
genes)

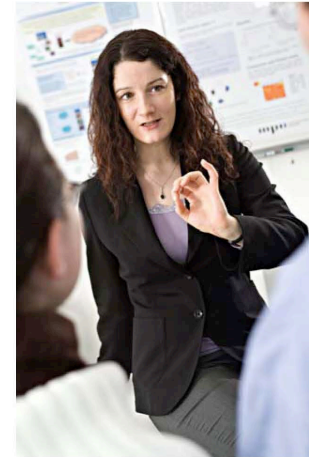


Alice C. McHardy



Research Interests

- Binning (taxonomic assignment) of metagenome sequence samples
- Methods for detecting sites/areas under selection in populations and adaptation to habitats
- Methods for functional protein characterization



Folker

- Comparative metagenomics
- Metagenome assembly
- Annotation
- Scaling up analysis
- “standards” // metadata



Christopher Quince

School of Engineering – University of Glasgow
quince@civil.gla.ac.uk

Research Interests

- Removing noise from amplicon data to allow accurate determination of community structure - PyroNoise
- Accounting for sampling effects to determine true diversities
- Stochastic models of microbial community assembly and function
- Integration of genomics data into community models
- Aim: improved application of microbial communities in environmental engineering
 - microbial fuel cells
 - water treatment
 - bioremediation
- To achieve this use ideas from Bayesian statistics, HPC, machine learning and mathematical ecology



With son Edward on summit of mighty Ben A'an (1488ft)

Jeroen Raes

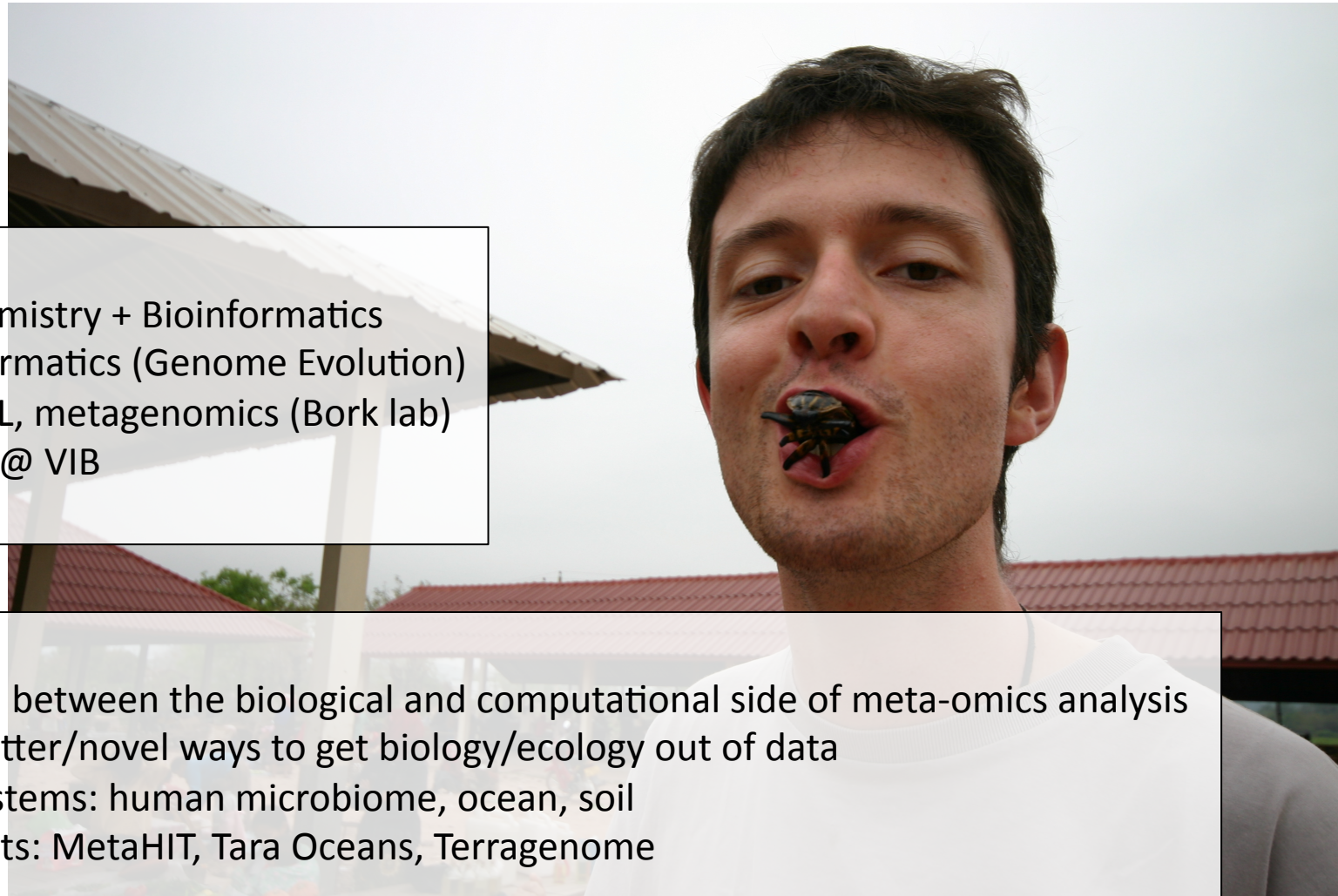
VIB/University of Brussels, Belgium

Background:

- MSc in Biochemistry + Bioinformatics
- PhD in Bioinformatics (Genome Evolution)
- PostDoc: EMBL, metagenomics (Bork lab)
- Now: Prof/GL @ VIB

Interests:

- Somewhere in between the biological and computational side of meta-omics analysis
- Looking for better/novel ways to get biology/ecology out of data
- Various ecosystems: human microbiome, ocean, soil
- Various projects: MetaHIT, Tara Oceans, Terragenome



Rick

Alex Sczyrba

DOE Joint Genome Institute

Project:

Large-scale discovery of biomass degrading genes and genomes from cow rumen

- 260 Gb Illumina data
- k-mer based quality filter
- Nucleating Assembly
- de-novo assembly
- >20,000 glycosyl hydrolases identified
- ~10 draft genomes assembled

Interests:

- Metagenome Assembly
- k-mer based data analysis
- cloud computing



Expertise :

community ecology
ecological theory
aquatic microbial ecology

Learning :

soils and host-associated
microbial communities

Research interests :

Temporal structuring of microbial
communities

Defining a “core” microbiome beyond
shared membership

Developing cross-ecosystem principles

Does community composition matter?

Ashley Shade

Microbial ecologist
Handelsman lab
Yale University

